

## **Study of the Feasibility of Utilizing Comparison-group Designs for Evaluating the North Carolina More at Four Pre-kindergarten Program**

Ellen Peisner-Feinberg, Ph.D.  
and Jennifer Schaaf, Ph.D.

FPG Child Development Research Institute  
The University of North Carolina at Chapel Hill

December 2008

# **Study of the Feasibility of Utilizing Comparison-group Designs for Evaluating the North Carolina More at Four Pre-kindergarten Program**

The North Carolina More at Four Pre-kindergarten Program was initiated in 2002 to provide high quality preschool experiences to at-risk children in North Carolina, especially those not previously served in preschool, to help them be better prepared for and more successful in school. Annual external evaluations have been conducted each year and have provided information about the growth in school readiness and academic skills of More at Four participants during their pre-k and kindergarten years. Children's performance has been compared to national standards, differences in performance have been examined for subgroups of children on the basis of levels of risk, and the influence of the quality of instructional practices in both pre-k and kindergarten have been studied, but no comparison group of similar North Carolina children has heretofore been examined.

This report considers six possible comparison-group designs for an evaluation of the effectiveness of the North Carolina More at Four Program:

- 1) Experimental Design: Random assignment of More at Four applicants to treatment or comparison groups
- 2) Waitlist Comparison: Preschool comparison of children enrolled in More at Four and on the waitlist
- 3) Program Standards Comparison: Preschool comparison of children who attend More at Four and who attend other programs not meeting the More at Four standards
- 4) Kindergarten Comparison: Comparison of children in kindergarten who did and did not attend More at Four the previous year
- 5) Regression Discontinuity Design (RDD): Comparison of More at Four graduates and new enrollees
- 6) Third-grade Performance: Comparisons of More at Four and non-More at Four attendees on EOG scores

An experimental design using random assignment is described for comparison to the subsequent quasi-experimental designs. Although such a design is desirable because of its ability to allow causal conclusions to be drawn, the technical feasibility of this design often precludes its usage. In cases where an experimental design cannot be used, there are several types of quasi-experimental designs that can provide useful information about the impact of the program. One of the most important determinants of the quality of the data from any quasi-experimental study is the similarity of the chosen comparison group to the target group (in this

case, More at Four attendees). The well-delineated eligibility requirements of the More at Four Program would facilitate the identification of an appropriate comparison group, although there would still be logistical challenges to overcome for each of these designs in recruiting and maintaining such a sample, as well as in gathering the necessary data for evaluating program effectiveness.

Each study design utilizes a different method for determining a comparison group against which to measure the performance of More at Four participants. All of the study designs that are considered here are scientifically valid, although each would have different challenges in obtaining a valid comparison group and around various factors which affect the generalizability of the results. In some designs, there also may be ethical issues to consider around the potential denial of services. For each of these designs, different decisions would need to be made about the scope of the research, including the type of outcome measures selected, the time frame for gathering data, the size of the sample and the selection process for choosing the specific sample (e.g., how many and how to select the classrooms, sites, counties, etc.), all of which would affect the feasibility and costs of conducting the study. For each study design, an overview of the critical design features is presented, followed by information about the participants, procedures, and outcome variables that would be utilized, the generalizability of the results based on such a design, issues around the logistical feasibility of conducting such a study for the More at Four Program, and the relative costs. Following these study descriptions, a summary chart (Table 1) provides comparisons across the various designs on these components.

## **Experimental Design: Random Assignment of More at Four Applicants to Treatment or Comparison Groups**

### **Design Overview**

A true experimental design is described as the most powerful approach to examining the impact of the More at Four Program on children's development. In contrast to the quasi-experimental designs detailed below, this type of design allows causal statements about the effects of the More at Four Program to be made with confidence. However, the ethical and logistical issues surrounding this type of design make it much less feasible for program evaluation than the quasi-experimental designs described subsequently. This experimental design would involve randomly assigning eligible More at Four applicants to either enter the More at Four Program (treatment group) or to not enter the program (comparison group), and would compare the development of the two groups over the course of their pre-k year. Individuals assigned to the comparison group would not be served in More at Four classrooms (and preferably not in sites operating More at Four classrooms), but may choose to utilize other types of early childhood educational programs or not. This type of design would require a large pool of eligible applicants, so that similar numbers of children could be assigned through a random selection process to the treatment or comparison groups. The number of applicants would need to be much larger than the number of slots available in the program to make random assignment possible. This type of design has been used in the past in classic research

studies such as the Abecedarian Study (Campbell & Ramey, 1994; Ramey & Campbell, 1984) and the Perry Preschool Project (Schweinhart, Montie, Xiang, Barnett, Belfield, & Nores, 2005; Weikart, Bond, & McNeil, 1978), both of which randomly assigned children to either receive intensive preschool experiences or not, and then longitudinally measured children's outcomes in many different areas over time.

This type of design carries some ethical implications, however, primarily around the decision to deny program services to one group for research purposes. These concerns are especially warranted if efforts need to be made to create a larger applicant pool for the research and/or to artificially restrict the number of available More at Four slots to allow for random assignment. If the naturally available pool of eligible program applicants is substantially larger than the number of slots available in the program, then these ethical concerns are reduced. In addition, the question of what sort of preschool experience the comparison group will have is a relevant concern. If applicants selected into the comparison group choose to attend a different pre-k program offering similar experiences, then the effects of More at Four are likely to appear weaker than they would if children in the comparison group attended no pre-k program or a program with lower standards than More at Four.

### **Participants**

All participants in the study would have applied and been determined to be eligible for the More at Four Program. Children would be randomly assigned to either the target group (those who are enrolled in More at Four) or the comparison group (those who are not enrolled in More at Four). Past research using this design has typically included a total sample size of approximately 100 children, although these studies were initiated 30-40 years ago at a time when other types of preschool programs were less available, so the comparison groups tended to have no formal pre-k experiences (Campbell & Ramey, 1994; Ramey & Campbell, 1984; Schweinhart et al., 2005; Weikart et al., 1978). These conditions allowed for greater distinctions between the treatment group which received the preschool intervention and the comparison group which generally had no preschool experience. However, given that most 4-year-olds attend some type of preschool program nowadays, a larger sample would likely be needed for examination of the More at Four Program since it would be expected that the majority of the comparison group would seek some other type of pre-k program.

### **Procedures**

Children would be recruited into the study after their More at Four applications had been reviewed and their eligibility for the More at Four Program had been established. Random assignment of children to either the target or the comparison group would follow eligibility verification. The same research data would be gathered on both groups of children, such as individual assessments of key school readiness skills (e.g., language or cognitive skills, teacher ratings of behavior). Two data collection points would be needed, one at the beginning of the pre-k year and one at the end of the pre-k year to measure the effects of program participation.

### **Outcome Variables**

The outcome variables would be selected to best represent key areas of interest, such as

language, math, general cognition, and behavioral performance. Child assessment measures would need to be age-appropriate, reliable and valid, and have content relevant to the goals of the More at Four Program, school readiness and school success.

### **Generalizability**

The generalizability of this study would be high. The two groups (target and comparison) would be equivalent in terms of eligibility characteristics for More at Four and parental motivation to seek out the program. An experimental design such as this would allow causal statements about the effects of More at Four to be made with confidence.

### **Logistical Feasibility**

The ease of conducting this type of study would be very low compared to other designs. This type of study would be more difficult to conduct in terms of program buy-in (finding programs willing to engage in a random assignment process for making selection decisions), sample recruitment (finding programs with enough applicants so that a sufficient comparison group would be available as well as recruiting sufficient numbers for the research), sample maintenance (the effort and incentives required to track and maintain participation in the comparison group), and data collection (the comparison group is likely to be spread out over numerous different sites, including different types of preschool settings and homes and would require permission from numerous care providers as well as parents for data collection).

### **Costs**

The costs of this type of study design would be higher relative to other proposed study design types. It would likely require substantial effort to locate programs with adequate numbers of applicants as well as to track and maintain the sample, especially for comparison group participants. While children in the More at Four classrooms (treatment group) would be easier to track and therefore less costly to assess, the comparison group would require greater effort to track, would likely be spread out in many more different settings, and would need to be provided with greater incentives for participating in the study, thereby generating greater costs for the research.

## **Quasi-experimental Designs**

### **Waitlist Comparison: Preschool Comparison of Children Enrolled in More at Four and on the Waitlist**

#### **Design Overview**

This quasi-experimental design would involve comparing children selected for enrollment by the More at Four Program and eligible applicants who were not selected for enrollment (i.e., those on the waitlist) during the pre-k year. The target group of children would be those who applied to the More at Four Program and were selected by the program to receive a slot. The comparison group would be eligible children who applied to the program, but did not receive a slot and were therefore unable to participate in the program. The development of these two groups of children would be assessed and compared over the course of the More at Four program year, in order to examine the achievement and growth rates for the two groups. Similar designs have been used in past research such as an early preschool investigation in New York (Irvine, 1982). The New York study compared target children who enrolled in a preschool program to a group of children who had applied, but remained on a waitlist due to a shortage of available spaces. For this design to be implemented to investigate More at Four, the number of More at Four applicants would have to be significantly larger than the number of available slots. This necessary situation has not been met in previous years of the More at Four Program.

There are also some potential concerns around the generalizability and scientific rigor of this design. The manner for choosing which applicants are selected to receive the More at Four slots could have significant effects in this regard. For instance, if the most needy applicants were selected to receive the More at Four slots (as is often the practice), this would generate a systematic difference between the target and comparison groups that would interfere with the study's ability to detect the effects of program participation (i.e., children at greater risk receiving the program would be compared to children at lesser risk but not receiving the program). Moreover, this type of research design would be jeopardized by the subsequent enrollment of comparison group participants in the More at Four Program if slots later become available (due to withdrawal of children from the program and/or expansion of the available number of slots). There are also some potential ethical issues around the denial of services with this type of design, similar to those for the experimental design. Such issues would arise in the case where more applicants than the More at Four Program can enroll are recruited solely to provide the necessary participants for research or where the number of available slots is restricted to provide sufficient numbers on the waitlist.

#### **Participants**

The target group would be children who were selected into the More at Four Program, and the comparison group would be children who applied but were not selected for enrollment by the More at Four Program. These two groups would be scientifically comparable in important

ways, including that both would have met all the eligibility requirements for More at Four, and they would also be similar on the less tangible, but meaningful, factor of parental motivation to seek out and apply for the program. It would be beneficial scientifically to include programs where these selection decisions are not based on characteristics related to children's performance on the study measures, such as risk levels, so that the comparison and target groups are more similar. It would also be important to gather information on the demographic characteristics of both groups, to determine whether they are different in any ways that also relate to the measurement of effects of participation in More at Four and to account for any differences as best as possible in the analysis of the results. The target and comparison groups would need to be of approximately equal size, and in previous years of the More at Four Program there have not been sufficient numbers of applicants to support this type of research design. Although past research has used sample sizes of several thousand (e.g., the 1982 Irvine study had a sample size of 5,000), it is likely that a substantially smaller sample size could still detect significant differences in the effects of participation in the More at Four Program.

### **Procedures**

Children would be recruited into the study after their More at Four applications have been reviewed and their eligibility for the More at Four Program has been established. The selection decisions about whether children are enrolled in More at Four or not would be made by the programs, as usual. The same research data would be gathered on both groups of children, such as individual assessments of key school readiness skills (e.g., language or cognitive skills, teacher ratings of behavior). Two data collection points would be needed, one at the beginning of the pre-k year and one at the end of the pre-k year to measure the effects of program participation.

### **Outcome Variables**

The outcome variables would be selected to best represent key areas of interest, such as language, math, general cognition, and behavioral performance. Child assessment measures would need to be age-appropriate, reliable and valid, and have content relevant to the goals of the More at Four Program, school readiness and school success.

### **Generalizability**

The generalizability of this study would be medium to high because the target group and the comparison group would be equivalent in terms of program eligibility as well as parental motivation to apply to More at Four. However, depending on how the program selects participants, there could be important differences in the characteristics of these two groups that could be detrimental to the ability to detect the effects of program participation.

### **Logistical Feasibility**

The ease of conducting this type of study would be low compared to other designs. This type of study would be more difficult to conduct in terms of sample recruitment (finding programs with enough applicants so that a sufficient comparison group would be available as well as recruiting sufficient numbers for the research), sample maintenance (the effort and incentives

required to track and maintain participation in the comparison group), and data collection (the comparison group is likely to be spread out over numerous different sites, including different types of preschool settings and homes and would require permission from numerous care providers as well as parents for data collection).

### **Costs**

The costs of this type of study design would be higher relative to other proposed study design types, for reasons similar to those of the experimental design. It would likely require substantial effort to locate programs with adequate numbers of applicants as well as to track and maintain the sample, especially for comparison group participants. While children in the More at Four classrooms (treatment group) would be easier to track and therefore less costly to assess, the comparison group would require greater effort to track, would likely be spread out in many more different settings, and would need to be provided with greater incentives for participating in the study, thereby generating greater costs for the research.

## **Program Standards Comparison: Preschool Comparison of Children Who Attend More at Four and Children Who Attend Other Programs Not Meeting the More at Four Standards**

### **Design Overview**

This quasi-experimental design would involve comparisons during the pre-k year of children who attend More at Four with similar children who attend other preschool programs that do not meet the standards of More at Four. The target group of children would be comprised of a random sample of those participating in the More at Four Program (selected at the county and/or classroom level). The comparison group would be selected from other preschool programs (in the same locales) that do not meet the standards of More at Four, with a matched sample of children who meet the eligibility characteristics for More at Four. Accordingly, the target and comparison groups would be children with similar characteristics, but with the difference being the type of pre-k program their families chose. The development of these two groups of children would be assessed and compared over the course of their pre-k year, in order to examine the achievement and growth rates for the two groups. This type of design is comparable to past studies that have included groups of children receiving preschool or other school interventions compared to similar groups of children enrolled in other sites that did not offer the same interventions (Reynolds, 1994; Starkey, Klein, & Wakeley, 2004).

### **Participants**

The target group of children would be children attending randomly-selected More at Four classrooms (the sample could be further stratified by county or other characteristics of interest). The comparison group would be selected from other preschool programs in the same locales which do not meet the standards of More at Four, but which serve children with similar characteristics based on eligibility for More at Four (e.g., age, poverty status, other risk factors). The programs children attend would be selected by the families; the research samples would be

selected from those already attending the sampled programs, both for the More at Four and the comparison programs. Past research using this design has commonly included sample sizes of 75-100 children per group, with total samples ranging from approximately 160 (Starkey et al., 2004) to 1,100 (Reynolds, 1994).

### **Procedures**

Classrooms from two groups of programs would be selected—More at Four Programs and other pre-k programs that do not meet the standards of More at Four, but who are serving similar groups of 4-year-olds. The More at Four programs and comparison programs would be selected from the same locales, to minimize other potential differences between the groups related to geographic region. The comparison children recruited to participate in the study would be selected to be similar to the More at Four children in the study, based on the eligibility characteristics for More at Four (e.g., age, poverty status, other risk factors). The same research data would be gathered on both groups of children, such as individual assessments of key school readiness skills (e.g., language or cognitive skills, teacher ratings of behavior). Two data collection points would be needed, one at the beginning of the pre-k year and one at the end of the pre-k year to measure the effects of program participation.

### **Outcome Variables**

The outcome variables would be selected to best represent key areas of interest, such as language, math, general cognition, and behavioral performance. Child assessment measures would need to be age-appropriate, reliable and valid, and have content relevant to the goals of the More at Four Program, school readiness and school success.

### **Generalizability**

The generalizability of this study design would be medium to high. The two types of programs would be readily distinguishable based on program standards. The comparison group and target groups would be matched on a number of child characteristics related to eligibility for More at Four. However, the two groups would differ in a key way in terms of parental motivation to enroll their children in More at Four.

### **Logistical Feasibility**

The ease of conducting this study would be medium. It is likely that the target and comparison groups would each include similar numbers of programs, classrooms, and children (although the number of sites could be higher in the comparison group), with similar effort required for data collection in both groups. It would require greater effort to obtain the comparison group, however, in terms of sample recruitment (finding More at Four and non-More at Four programs in the same locales that have different program standards but are serving similar children), and sample maintenance (larger incentives may be required to generate participation at the program and child level for the comparison group).

### **Costs**

The costs for this type of study design would be medium compared with other proposed study designs. Children in both the target and comparison groups would be selected from a limited number of classrooms, facilitating sample recruitment, maintenance, and data collection. However, greater effort would be required to find and recruit eligible children from the comparison classrooms; not all children in the selected classrooms would be eligible for the study, given the necessity of selecting a matched sample of children (i.e., those who match the eligibility characteristics of More at Four). In addition, it may require greater effort and larger incentives to recruit comparison programs and children in those classrooms, given that they are not attending the program of focus (More at Four).

## **Kindergarten Comparison: Comparison of Children in Kindergarten Who Did and Did Not Attend More at Four the Previous Year**

### **Design Overview**

This quasi-experimental design would involve comparing two groups of children during their kindergarten year—those who attended More at Four the previous year and similar children in the same kindergarten classrooms who did not attend More at Four. The target group of children would be those who attended More at Four the previous year. The comparison group would consist of children who had different preschool experiences (i.e., those who attended a different preschool program or no program during the previous year). Within each kindergarten classroom, children from both groups (target and comparison) would be included. The comparison group would need to be carefully matched on all of the More at Four eligibility requirements, and their preschool experiences would need to be accurately determined. Research studies have found that accurate information about prior experiences, even during the previous year, is not always easily obtainable from teachers or parents, although these would be the most feasible sources. To examine only the question of how the school readiness of children who participate in More at Four differs from other children of similar background (but with different preschool experiences), only one time point at the beginning of the kindergarten year would be needed for data collection. For more complete information about the learning trajectory of the two groups, two data collection points would be needed, one at the beginning and one at the end of the kindergarten year. This would allow a comparison of academic achievement and the size of the achievement gap, if any, between the target and comparison groups during the first year of school. The second data collection point would also allow the learning trajectories of the two groups to be compared, and would allow inferences about how growth might continue as well as when and if a gap between the two groups might be expected to close. This type of study design has been utilized to examine other programs, such as comparisons of children who attended centers receiving Smart Start quality improvement funds versus those who attended other centers (Maxwell, Bryant, & Miller-Johnson, 1999).

### **Participants**

A representative sample of kindergarten children who had participated in More at Four the previous year would be chosen (using a random or stratified random procedure). Then a

sample of children from the same kindergarten classes who are matched in terms of eligibility characteristics for More at Four (e.g., age, poverty status, other risk factors), but who did not participate in More at Four or attend programs funded by More at Four would be chosen. It would be beneficial to sample from counties and schools that are likely to have multiple potential target and comparison children in the same kindergarten classrooms. It would be important to accurately ascertain the preschool experiences of the comparison sample by teacher and/or parent report, which are the most feasible sources but are not always accurate. Past research using this design has included total sample sizes of approximately 500 children (Maxwell, Bryant, & Miller-Johnson, 1999).

### **Procedures**

The same data would be gathered on both groups of children, such as individual assessments (e.g., language or cognitive skills, teacher ratings of behavior). To determine only how the school readiness of children who experienced More at Four compares to children who attended other programs, only one data collection point would be needed at the beginning of the kindergarten year. To determine school readiness, rate of growth, and the academic achievement of the two groups after a year of kindergarten, two data collection points would be needed at the beginning and at the end of the kindergarten year.

### **Outcome Variables**

The outcome variables would be selected to best represent key areas of interest, such as language, math, general cognition, and behavioral performance. Child assessment measures would need to be age-appropriate, reliable and valid, and have content relevant to the goals of the More at Four Program, school readiness and school success.

### **Generalizability**

The generalizability of this design would be medium. The target and comparison groups would be matched on the eligibility characteristics for More at Four as much as possible, thereby reducing differences between the groups that could affect their outcomes but are not related to pre-k program participation. However, the two groups would still differ in a key way in terms of parental motivation to enroll their children in More at Four. Moreover, it would be important to insure that reliable methods were used to gather information about children's pre-k experiences during the previous year, especially those in the comparison group (since other data are available for tracking More at Four participants), to insure that appropriate comparisons are made. Depending on the numbers in different groups and the accuracy and availability of detailed information about children's pre-k experiences the previous year, it may be useful to make sub-group comparisons based on the similarity of the comparison programs to the More at Four programs as well as comparisons of children with no formal pre-k experience versus those in other types of pre-k programs versus More at Four.

### **Logistical Feasibility**

The ease of conducting this study would be medium to high compared to other designs. Children in both groups would be selected from the same kindergarten classrooms, facilitating

many of the study procedures. However, it may require substantial effort at the recruitment stage to find adequate numbers of kindergarten classrooms serving both children who attended More at Four and similar children who did not attend More at Four. In addition, study procedures would need to insure that accurate data are gathered about the pre-k experiences of these kindergarten children, either from teachers or parents.

### **Costs**

The costs for this study design would be low to medium compared to the other proposed types of study designs. Children in both the target and comparison groups would be located in the same kindergarten classes, reducing costs related to sample selection and recruitment and data collection compared to some other designs. However, it may require greater effort to locate a sufficient number of schools and kindergarten classrooms serving adequate numbers of children for both the target and comparison groups, and it would likely be necessary to limit the sample to selected counties or other geographic regions in the state.

## **Regression Discontinuity Design (RDD): Comparison of More at Four Graduates and New Enrollees**

### **Design Overview**

This quasi-experimental design would involve comparing children entering kindergarten who had just completed the More at Four Program (target group) with children who were about to enter the More at Four Program (comparison group). Both groups of children would be compared at the same time, based on one assessment point at the beginning of the school year; for the first group, this would be at entry into kindergarten and for the second group, at entry into pre-k. The two groups would be comparable in terms of their eligibility for and motivation to attend More at Four. The difference would be in terms of experience with the program; the target group would have participated in More at Four, while the comparison group would not yet have had the chance to participate (because they are just starting pre-k). The linear association between age and performance would be examined separately for the two groups; differences in these trajectories would then represent the effects of program participation. The idea behind this design is that the two groups of children would differ only in terms of their pre-k experience, but not in terms of other characteristics such as eligibility, demographic characteristics, or motivation to seek out and attend the pre-k program. In the truest form of RDD, the children in each group would be very similar in age, with children in the first group just making the birth date cutoff for eligibility for the program the previous year while the second group just missed the birth date cutoff and therefore was not eligible to attend until the subsequent year. That way, age differences between the two groups that are also related to their level of knowledge, skills, and development (an especially important consideration for young children) would be minimized. Because locating a large enough sample of children with such a restricted range of birth dates is not feasible, this type of design has been implemented based on a wider range of birth dates (e.g., children whose birth dates range within 6 months of the cutoff date). However, the wider the range, the less similar the two groups would be in

terms of age and age-related skills; for example, such ranges could result in 5-year-old-children in the target group being compared with 4-year-old children in the comparison group, which at those ages, represents a 20% difference in children's life experiences. This type of design has been used to evaluate state pre-kindergarten initiatives in single states, such as in Oklahoma (Gormley, Gayer, Phillips, & Dawson, 2005) as well as programs across several states including Michigan, New Jersey, Oklahoma, South Carolina, and West Virginia (Wong, Cook, Barnett & Jung, 2008).

### **Participants**

The target group would be children who had participated in More at Four the previous year, and are ready to begin kindergarten. The comparison group would consist of children who are just entering the More at Four Program. Children would be selected from the same classrooms and/or programs for both groups, with the target group having just attended the More at Four classroom the previous year and comparison group being the current enrollees. These two groups would be well-matched on eligibility for the More at Four Program and parental motivation to seek out and apply to the program. However, there would be some discrepancies in age between the two groups, which could be substantial depending on the feasibility of finding children in both groups with birth dates close to the cutoff. Past research using this design has typically enrolled approximately 3,000 to 5,000 participants (Gormley et al., 2005; Wong, et al., 2008), although it is likely that a substantially smaller sample size could still detect significant differences in the effects of participation in the More at Four Program.

### **Procedures**

The same data would be gathered on both groups of children, such as individual assessments (e.g., language or cognitive skills, teacher ratings of behavior). Only one data collection point would be needed, at the beginning of the school year (kindergarten for the target group and pre-k for the comparison group). It would be important to conduct the assessments as close to the beginning of the school year as possible, to maximize the differences in program experiences between the two groups; in other words, to insure that the comparison group has not had the pre-k program experience that the target group has received. Ideally, children would be assessed prior to the beginning of the school year, but it is not feasible procedurally to find and assess children outside of their kindergarten or pre-k classrooms.

### **Outcome Variables**

The outcome variables would be selected to best represent key areas of interest, such as language, math, general cognition, and behavioral performance. Measures would need to be age-appropriate (for both pre-k and kindergarten), reliable and valid, and have content relevant to goals of the More at Four program, school readiness and school success.

### **Generalizability**

The generalizability of this design would be medium to low, depending on how close in age the target and comparison groups are (the closer in age, the better the generalizability). This type of design also assumes that growth in the skills measured is linearly related to age; the extent to

which that assumption is true for the measures selected would also affect the generalizability. The target and comparison groups would be equivalent both in terms of eligibility characteristics for More at Four and parental motivation to enroll their children in More at Four.

### **Logistical Feasibility**

The ease of conducting this study would be low to medium compared to other designs. This design would require a relatively large sample size and substantial effort at the recruitment stage to find children in both groups from the same settings within a limited range of birth dates. In addition, it is likely that the target group would require more effort for data collection, given that they would be spread out over more classrooms and schools in kindergarten than the comparison group in More at Four. A high amount of effort would also be required for recruiting and training data collectors to insure that the data collection period occurs over a short window of time as early as possible in the school year, to maximize the differences between the two groups (and minimize the pre-k experiences of the comparison group).

### **Costs**

The costs for this study design would be medium compared to the other proposed types of study designs. Though only one point of data collection would be needed, it would need to occur during a short window of time with a relatively large number of participants, thus increasing costs related to training and hiring data collectors. Children in both the target and comparison groups would be in the same geographic locations, though not necessarily the same schools (only half the More at Four classrooms are located in school-based settings). A relatively large sample would be needed for this type of study, requiring higher costs in terms of sample selection and recruitment.

## **Third-grade Performance: Comparisons of More at Four and non-More at Four Attendees on Third-grade EOG Scores**

### **Design Overview**

This quasi-experimental design would involve school-age comparisons of children's academic performance, relying on extant data gathered by the schools, of children who attended More at Four during pre-k (target group) and similar children from the same schools or districts who did not attend More at Four (comparison group). This study design would utilize the North Carolina End-of-Grade Mathematics and Reading Comprehension—Grade 3 (EOG) test scores (or other extant academic performance data as available). Scores from children who participated in More at Four during pre-k would be compared to those of children from similar backgrounds who did not participate in More at Four to examine their longer-term academic performance. The comparison group would need to be matched as closely as possible to the target group in terms of eligibility characteristics for More at Four (e.g., poverty status, other risk factors), although the extant data available for these factors may be limited. Other studies have utilized third-grade test score data, such as the evaluation of Georgia's pre-k program,

although that study only examined differences in program participants and did not include a comparison group (Henry, Gordon, Henderson, & Ponder, 2001).

### **Participants**

Participants would be selected from all children with available third-grade EOG scores in a given year. Information collected by the More at Four Program at the state level would be used to document and track children who attended More at Four in pre-k. The comparison group would be selected from children with similar characteristics (such as poverty status based on eligibility for free or reduced-price lunch, IEP status) who attend the same schools and/or school districts. Past research studies have included approximately 2,000 children, although no comparison group was included (Henry et al., 2001).

### **Procedures**

Existing data from the third-grade math and reading EOGs would be examined, along with extant data about children's demographic characteristics. No new data collection would occur. EOG scores of students who had participated in More at Four during pre-k would be compared to scores of students who did not attend More at Four who were matched on demographic characteristics.

### **Outcome Variables**

The outcome measures would be comprised of the third-grade EOG scores for reading and math, representing children's academic performance. No additional outcome measures would be gathered.

### **Generalizability**

The generalizability of this study is low to medium. The target and comparison groups would be matched in terms of available demographic characteristics related to More at Four eligibility requirements (likely to be limited to eligibility for free or reduced-price lunch and IEP status), but information about the preschool experiences of the comparison group will not be available. It is likely that some children in the comparison group will have received other similar preschool experiences, such as attendance at a 5-star child care center, while others may have attended programs not meeting the standards of More at Four or had no formal pre-k program experience. These differences within the comparison group in preschool experiences will not be known, and such unknown variables will reduce the ability to draw conclusions about the effects of the More at Four Program. Also, the outcome variables that can be examined will be limited to the available extant test score data. Further, information about the instructional quality of children's subsequent school experiences from kindergarten through third grade will not be available to take into account, although differences in such experiences are likely to affect performance on these measures.

### **Logistical Feasibility**

The ease of conducting this study would be high compared to other proposed study design

types. Extant data provided by the state would be used for these analyses; thus there are no new data collection efforts involved. Assuming de-identified data could be utilized, no subject recruitment efforts would be required. The feasibility of obtaining adequate usable data is very good; the only difficulties likely to arise are in accurately identifying previous More at Four students, the availability of other demographic data needed to match the target and comparison samples (such as free or reduced-price lunch eligibility), the ability to link these data with the EOG data, and the ability to find adequate matched samples within schools and/or school districts.

### **Costs**

The costs for this type of study design are low compared to other proposed study designs, given that it relies on extant data. No sample recruitment or new data collection is needed, thereby reducing costs considerably compared to other types of designs.

**Table 1: Summary Comparison of Key Features of Different Study Designs**

Design Type	Study Groups: Target vs Comparison	Generalizability	Logistical Feasibility	Relative Costs
Experimental Design	MAF applicants randomly assigned to enroll in MAF <i>vs</i> Applicants not assigned to MAF	High	Very Low	Higher
Waitlist Comparison	MAF enrollees <i>vs</i> MAF applicants on waitlist (not enrolled in MAF)	Medium/High	Low	Higher
Program Standards Comparison	Children in MAF programs <i>vs</i> Children attending other programs not meeting MAF standards	Medium/High	Medium	Medium
Kindergarten Comparison	Kindergartners who participated in MAF in pre-k <i>vs</i> Kindergartners who did not participate in MAF	Medium	Medium/High	Low/Medium
Regression Discontinuity Design	Kindergarten graduates of MAF <i>vs</i> New pre-k enrollees in MAF	Medium/High	Low/Medium	Medium
Third-grade Performance	3 <sup>rd</sup> -graders who attended MAF <i>vs</i> 3 <sup>rd</sup> -graders who did not attend MAF	Low/Medium	High	Lower

## References

- Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, 65, 684-698.
- Gormley, W., Gayer, T., Phillips, D., & Dawson, B. (2005, November). The Effects of Universal Pre-K on Cognitive Development. *Developmental Psychology*, 41(6), 872-884.
- Irvine, D. J. (1982, March). Evaluation of the New York state experimental prekindergarten program. Paper presented at the meeting of the American Educational Research Association, New York, NY. (ERIC document reproduction Service No. ED 17980)
- Maxwell, K., Bryant, D., & Miller-Johnson, S. (1999). A six-county study of the effects of Smart Start child care on kindergarten entry skills. Chapel Hill, NC: FPG Child Development Institute. (ERIC Document Reproduction Service No. ED 433154)
- Ramey, C., & Campbell, F. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, 88(5), 515-523.
- Reynolds, A. J. (1994). Effects of a preschool plus follow-on intervention for children at risk. *Developmental Psychology*, 30, 787-804.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime Effects: The High/Scope Perry Preschool study through age 40. (Monographs of the High/Scope Educational Research Foundation, 14). Ypsilanti, MI: High/Scope Press.
- Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 19, 99-120.
- Weikart, D. P., Bond, J. T., & McNeil, J. T. (1978). The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade. Ypsilanti, MI: High/Scope Press.
- Wong, V., Cook, T., Barnett, W., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122-154.